Principal Component Analysis

Yuanzhen Shao

MA 26500

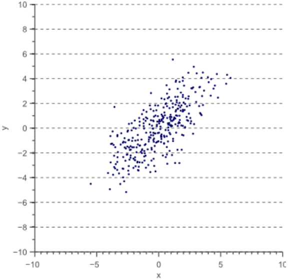
Data as points in $\mathbb{R}^n$

Assume that we have a collection of data

$$S = \{X_1 = \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1n} \end{bmatrix}, X_2 = \begin{bmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2n} \end{bmatrix}, \cdots, X_m = \begin{bmatrix} x_{m1} \\ x_{m2} \\ \vdots \\ x_{mn} \end{bmatrix}\}$$

in $\mathbb{R}^n$.

Figure: A collection data in $\mathbb{R}^n$



Data as points in $\mathbb{R}^n$

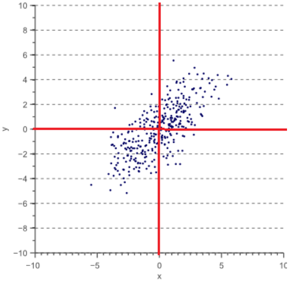
Without loss of generality, we may assume that the mean of this collection of data

$$\bar{X} = \frac{1}{m}\sum_{i=1}^{m} X_i = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Otherwise, we just translate these data to have 0 mean by looking at

$$\bar{S} = \{X_1 - \bar{X}, X_2 - \bar{X}, \cdots, X_m - \bar{X}\}.$$

Figure: A collection data in $\mathbb{R}^n$



Variance of Data

Question 1: How can we determine a subspace that S is close to?

Some examples: http://setosa.io/ev/principal-component-analysis/

Variance of Data

Question 1: How can we determine a subspace that S is close to?

Some examples: http://setosa.io/ev/principal-component-analysis/

Question 2: How to determine the direction representing the largest variance of S?

Variance of Data

Question 1: How can we determine a subspace that S is close to?

Some examples: http://setosa.io/ev/principal-component-analysis/

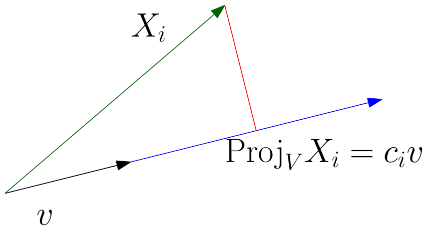
Question 2: How to determine the direction representing the largest variance of S?

Recall if $v \in \mathbb{R}^n$ is a unit vector, then the orthogonal projection of X_i on the direction given by v, i.e. span$\{v\}$, is

$$\mathrm{Proj}_V X_i = (X_i, v)v = c_i v,$$

that is, c_i is the coordinate of X_i in the direction of v.

Figure: Orthogonal projection



Variance of Data

Answer: the direction representing the largest variance of S is given by the unit vector v that can maximize

$$\sqrt{\sum_{i=1}^{m} c_i^2} = \sqrt{\sum_{i=1}^{m} (X_i, v)^2} \qquad \text{among all unit vector } v,$$

or equivalently to maximize

$$\sum_{i=1}^{m} c_i^2 = \sum_{i=1}^{m} (X_i, v)^2 \qquad \text{among all unit vector } v.$$

Optimization Problem

Question 3: How to find such v to this optimization problem?

Let

$$A_{m\times n} = \begin{bmatrix} X_1^T \\ X_2^T \\ \vdots \\ X_m^T \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

Optimization Problem

Question 3: How to find such v to this optimization problem?

Let

$$A_{m\times n} = \begin{bmatrix} X_1^T \\ X_2^T \\ \vdots \\ X_m^T \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

Recall $(X_i, v) = X_i^T v$. So

$$Av = \begin{bmatrix} X_1^T v \\ X_2^T v \\ \vdots \\ X_m^T v \end{bmatrix} = \begin{bmatrix} (X_1, v) \\ (X_2, v) \\ \vdots \\ (X_m, v) \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix}.$$

Optimization Problem

Question 3: How to find such v to this optimization problem?

Let

$$A_{m\times n} = \begin{bmatrix} X_1^T \\ X_2^T \\ \vdots \\ X_m^T \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

Recall $(X_i, v) = X_i^T v$. So

$$Av = \begin{bmatrix} X_1^T v \\ X_2^T v \\ \vdots \\ X_m^T v \end{bmatrix} = \begin{bmatrix} (X_1, v) \\ (X_2, v) \\ \vdots \\ (X_m, v) \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix}.$$

Therefore,

$$\sum_{i=1}^{m} c_i^2 = (Av, Av) = v^T A^T A v = v^T \underbrace{C}_{=A^T A} v$$

Optimization Problem

C is an $n \times n$ symmetric matrix (consider why), and thus is diagonalizable.

C is an $n \times n$ symmetric matrix (consider why), and thus is diagonalizable.

Assume that C has eigenvalues

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n.$$

Moreover, C has an orthonormal basis of eigenvectors

$$v_1, v_2, \cdots v_n \qquad \text{such that } Cv_i = \lambda_i v_i.$$

C is an $n \times n$ symmetric matrix (consider why), and thus is diagonalizable.

Assume that C has eigenvalues

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n.$$

Moreover, C has an orthonormal basis of eigenvectors

$$v_1, v_2, \cdots v_n \qquad \text{such that } Cv_i = \lambda_i v_i.$$

In particular,

$$v_i^T C v_i = \lambda_i v_i^T v_i = \lambda_i \|v_i\|^2 = \lambda_i.$$

C is an $n \times n$ symmetric matrix (consider why), and thus is diagonalizable.

Assume that C has eigenvalues

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n.$$

Moreover, C has an orthonormal basis of eigenvectors

$$v_1, v_2, \cdots v_n \qquad \text{such that } Cv_i = \lambda_i v_i.$$

In particular,

$$v_i^T C v_i = \lambda_i v_i^T v_i = \lambda_i \|v_i\|^2 = \lambda_i.$$

Claim: v_1 represents the direction of the largest variance of S.

Proof.

If $u = a_1 v_1 + a_2 v_2 \cdots + a_n v_n$ is a unit vector in $\mathbb{R}^n$, i.e.

$$1 = \|u\|^2 = a_1^2 + a_2^2 + \cdots + a_n^2, \qquad \text{(consider why?)}$$

then

$$\begin{aligned} u^T C u &= \underbrace{(a_1 v_1 + a_2 v_2 \cdots + a_n v_n)^T}_{u^T} \underbrace{(\lambda_1 a_1 v_1 + \lambda_2 a_2 v_2 \cdots + \lambda_2 a_n v_n)}_{Cu} \\ &= \lambda_1 a_1^2 + \lambda_2 a_2^2 + \cdots \lambda_n a_n^2 \\ &\leq \lambda_1 a_1^2 + \lambda_1 a_2^2 + \cdots \lambda_1 a_n^2 \\ &= \lambda_1 \underbrace{(a_1^2 + a_2^2 + \cdots + a_n^2)}_{=1} \\ &= \lambda_1. \end{aligned}$$

□

Optimization Problem

Thus,

$$v_1 = \text{ the direction of the largest variance of } S.$$

Thus,

$$v_1 = \text{ the direction of the largest variance of } S.$$

Similarly, we can show that

- $v_2 =$ the direction of the largest variance of S in $\text{span}\{v_1\}^\perp$

Thus,

$$v_1 = \text{ the direction of the largest variance of } S.$$

Similarly, we can show that

- $v_2 =$ the direction of the largest variance of S in span$\{v_1\}^\perp$
- $v_3 =$ the direction of the largest variance of S in span$\{v_1, v_2\}^\perp$

Optimization Problem

Thus,

$$v_1 = \text{ the direction of the largest variance of } S.$$

Similarly, we can show that

- $v_2 =$ the direction of the largest variance of S in $\text{span}\{v_1\}^\perp$
- $v_3 =$ the direction of the largest variance of S in $\text{span}\{v_1, v_2\}^\perp$
- etc.

Thus,

$$v_1 = \text{ the direction of the largest variance of } S.$$

Similarly, we can show that

- $v_2 =$ the direction of the largest variance of S in $\text{span}\{v_1\}^\perp$
- $v_3 =$ the direction of the largest variance of S in $\text{span}\{v_1, v_2\}^\perp$
- etc.

In the end, we can just drop the directions corresponding to very small eigenvalues of C.

PCA in digital images: an example by Václav Hlaváč

- Let us consider a 321×261 image.

- Such an image can be considered as a vector in $\mathbb{R}^n$ with $n = 321 \times 261 = 83781$.

What if we have 32 instances of images?

PCA in digital images: an example by Václav Hlaváč

- Using PCA method, we can determine a four-dimensional subspace W in $\mathbb{R}^n$ such that all 32 images are close to W.

PCA in digital images: an example by Václav Hlaváč

- Using PCA method, we can determine a four-dimensional subspace W in $\mathbb{R}^n$ such that all 32 images are close to W.
- We can find four basis vectors for W, which can be displayed as images:

PCA in digital images: an example by Václav Hlaváč

- Using PCA method, we can determine a four-dimensional subspace W in $\mathbb{R}^n$ such that all 32 images are close to W.
- We can find four basis vectors for W, which can be displayed as images:

- We can reconstruct all 32 images by using linear combinations of these four basis images, e.g.

where $q_1 = 0.078$, $q_2 = 0.062$, $q_3 = -0.182$, $q_4 = 0.179$.

Reconstruction fidelity, 4 components

References

- Václav Hlaváč, Principal Component Analysis Application to images
- http://setosa.io/ev/principal-component-analysis/
- http://www.visiondummy.com/2014/04/geometric-interpretation-covariance-matrix/